(2)

SIMULATING THE IMPACT OF MPTS TRADE-OFF DECISIONS BY APPLICATION OF THE ISOPERFORMANCE METHODOLOGY

FINAL REPORT

Submitted to:

Commanding Officer
Naval Training Systems Center
12350 Research Parkway
Attn: Mr. Al Galbavy, PM Trade, Code HEL
Orlando, Florida 32826-3275

Submitted by:

Essex Corporation
1040 Woodcock Road
Orlando, Florida 32803

29 July 1992

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution unlimited.
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Essex Corporation; 6b. OFFICE SYMBOL (If applicable)	7a. NAME OF MONITORING ORGANIZATION Naval Training Systems Center
6c. ADDRESS (City, State, and ZIP Code) 1040 Woodcock Road, Suite 227 Orlando, FL 32803	7b. ADDRESS (City, State, and ZIP Code) 12350 Research Parkway Orlando, Fl. 32826-3275
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Naval Training Systems Center; 8b. OFFICE SYMBOL (If applicable)	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER N61339-92-C-0037
8c. ADDRESS (City, State, and ZIP Code)	10. SOURCE OF FUNDING NUMBERS: PROGRAM ELEMENT NO. / PROJECT NO. 2611P7 / TASK NO. / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Simulating the Impact of MPTS Trade-off Decisions by Application of the Isoperformance Methodology

12. PERSONAL AUTHOR(S)
Robert S. Kennedy and Marshall B. Jones

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM 1/29/92 TO 7/29/92	92/07/29	28

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	isoperformance, trade-offs, cost-effectiveness, MANPRINT, personnel, training, and equipment (MPT)

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Isoperformance is a trade-off methodology. Given a desired or good-enough level of performance, it indicates all combinations of performance determinants that will produce that level of performance. The traded-off determinants may be personnel aptitude and training, two equipment variants, two kinds of training (for example, simulator and field training or simulator alone and distributed interactive simulation), or any other pair of determinants. Visual presentation of isoperformance outcomes becomes a problem, but extensions to three or more determinants are also possible and, analytically, uncomplicated. The present report begins with a discussion of isoperformance methodology in relation to HARDMAN III. Very briefly, isoperformance methods and HARDMAN III are complementary. HARDMAN III determines manpower, personnel, and training constraints within which a design solution adequate to produce the required level of system performance must be found. Isoperformance eventuates in isoperformance curves (trade-off functions) adequate to produce the

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Mr. Al Galbavy	22b. TELEPHONE (Include Area Code) (407) 380-4267; 22c. OFFICE SYMBOL PM Trade, Code HEL

DD FORM 1473, 84 MAR — 83 APR edition may be used until exhausted. All other editions are obsolete.

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

19. Abstract (continued)

required level of system performance. If one starts with HARDMAN III, isoperformance methodology fills in the solution space within the HARDMAN constraints. If one starts with isoperformance curves, HARDMAN constraints can then be imposed to exclude unacceptable solutions. The result (a set of acceptable determinant combinations) is the same in either case. A final design solution can then be determined by cost-benefit analysis, which isoperformance curves allow. Following this discussion, the report takes up the case of trade-offs either from simulator training to field training or from one training simulator to another. A showing is made that Roscoe's treatment of this case (transfer effectiveness ratios) is derivative from isoperformance theory. Finally, a series of real-world isoperformance curves are constructed from data archived at the Navy Training and Performance Data Center (TPDC). The curves concern tank-maintenance mechanics and all four involve the trade-off between aptitude and training plus time-in-unit. The results indicate that low-aptitude mechanics (the bottom third) require as much as six months longer to become proficient than do average or high-aptitude mechanics (the upper two thirds).

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

INTRODUCTION

Isoperformance is a general methodology and applies in various ways in many trade-off situations (Kennedy, Jones, & Baltzley, 1989). We are aware, however, that the Army's STRICOM Command has particular interests in HARDMAN III and in the Army's Combined Arms Tactical Training (CATT) and component systems. Rather, therefore, than develop isoperformance as a general methodology, we will focus on what we take to be STRICOM's key concerns.

Altogether this report consists of four sections after this Introduction. The first of these sections will be a discussion of isoperformance methodology in relation to HARDMAN III. The section after that will take up trade-offs between simulator and field time for collective training. Among other things, it will include a showing that isoperformance theory subsumes transfer effectiveness ratios (Roscoe, 1971) as a special case.

The third following section presents the results of our Phase I effort. Very briefly, we were able to construct isoperformance curves for four tank-maintenance MOSs from data archived at the Training and Performance Data Center (TPDC). This result is significant not only for its demonstration that real-world isoperformance curves can be constructed from available data, but also because what was done for these four MOSs can also be done for most MOSs in all three services, provided only that the numbers of cases archived at TPDC are sufficient to give reliable results and that the data for them are sufficiently complete.

The last section is a brief discussion of the background and need for trade-off analysis, especially as it relates to MANPRINT, MOS restructuring, and other problems of human systems integration.

Additionally, a computer program is included with this report in order to illustrate the isoperformance methodology.

ISOPERFORMANCE METHODOLOGY

The current importance of trade-off methodology derives from several sources. One of them is the advent of SIMNET and distributed interactive simulation. The prospect of rapidly reconfigurable simulations creates possibilities for training that were out of the question only a few years ago. At present, even though we know that soldiers in training devices learn at different rates and have different aptitudes, most current military training devices systems offer a lock-step syllabus and most simulators offer the same scene content to all. Few simulators permit us to customize training for different trainees. Advances in computer simulation technology and, in the longer term, virtual reality will cause us to rethink this position since it will be not only possible but feasible to rapidly reconfigure a simulator to accommodate differences in trainee background, aptitudes, and experience. MANPRINT and HARDMAN are two other recent developments that place a premium on trade-off methods.

Cost-effectiveness methods (and isoperformance is a cost-effectiveness as well as a trade-off methodology) may proceed in either of two general ways. The more familiar is to fix costs and maximize effectiveness. One gets, as the popular phrase puts it, "the biggest bang for the buck." The alternative procedure is to fix effectiveness and minimize costs of health, safety, personnel, training, and equipment -- to get "the same bang in the least costly and most expeditious way." This latter approach leads naturally to trade-offs among cost factors and is the approach taken by isoperformance methodology (Jones, Kennedy, & Kuntz, 1987; Jones, Kennedy, Kuntz, & Baltzley, 1987; Kennedy, Jones, & Baltzley, 1988).

The heart of this methodology is the isoperformance curve. With respect to aptitude levels and training times, such a curve looks like the one given in Figure 1. The Y-axis is aptitude as measured, for example, by the Armed Forces Qualification Test (AFQT). The X-axis is training time in weeks. The job might be engine mechanic for a new helicopter. The curve drawn is for 80% proficient. That is, any point on the curve (any of the indicated combinations of aptitude level and training time) will produce students 80% of whom are proficient at the job. Thus, if one has high-aptitude people (for example, mental categories 1 and 2 on the AFQT), 80% proficient can be reached in roughly six weeks. With lower aptitude people, more training time is needed and for some aptitude levels (mental category 4 on the AFQT, perhaps) no amount of training time up to the maximum considered will suffice to produce students 80% of whom are proficient.

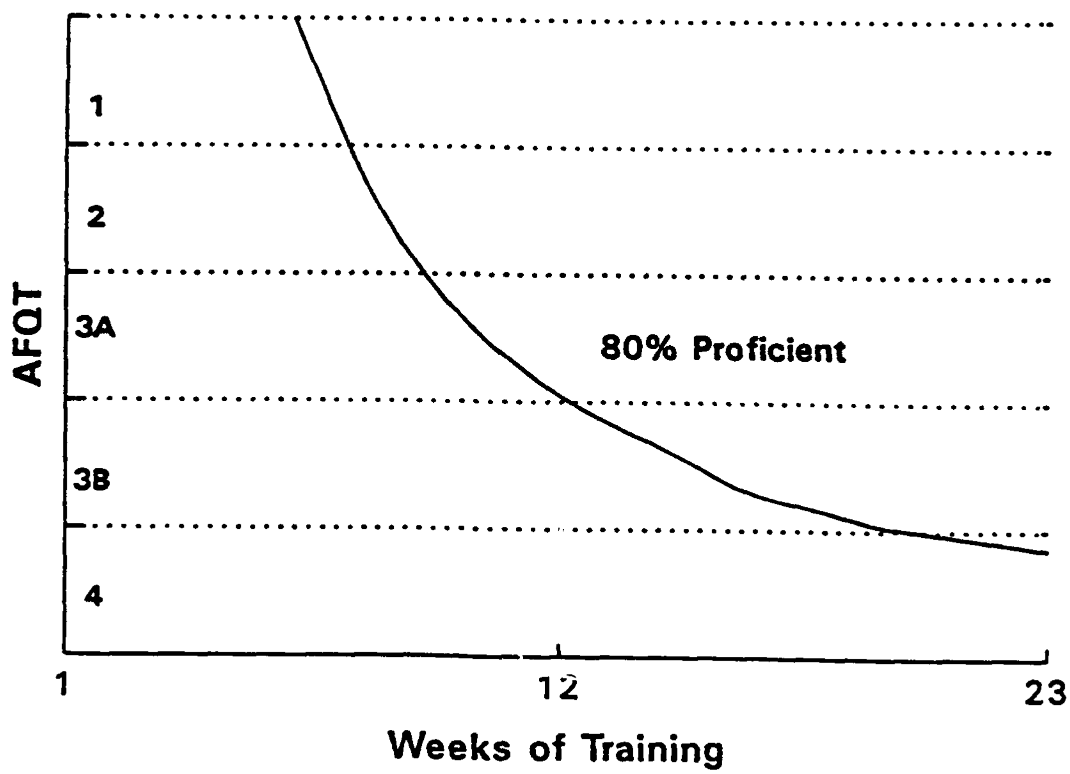

Figure 1. An isoperformance curve for 80% proficient.

Isoperformance curves come in families. A separate and distinct isoperformance curve exists for every level of performance that one specifies (Kennedy, Jones, & Baltzley, 1988; Jones & Jones Research Associates & Essex Corporation, 1987). Thus, if one were to specify 50% proficient, for example, one would get another curve than the one that appears in Figure 1. Figure 2 presents such a curve, along with the one in Figure 1. Note that the second curve lies to the left and down from the first curve presented. It takes less time to train the students to the lower level of performance. Or, in the alternative, for the same amount of training time one can attain the lower level of proficiency with lower aptitude people.

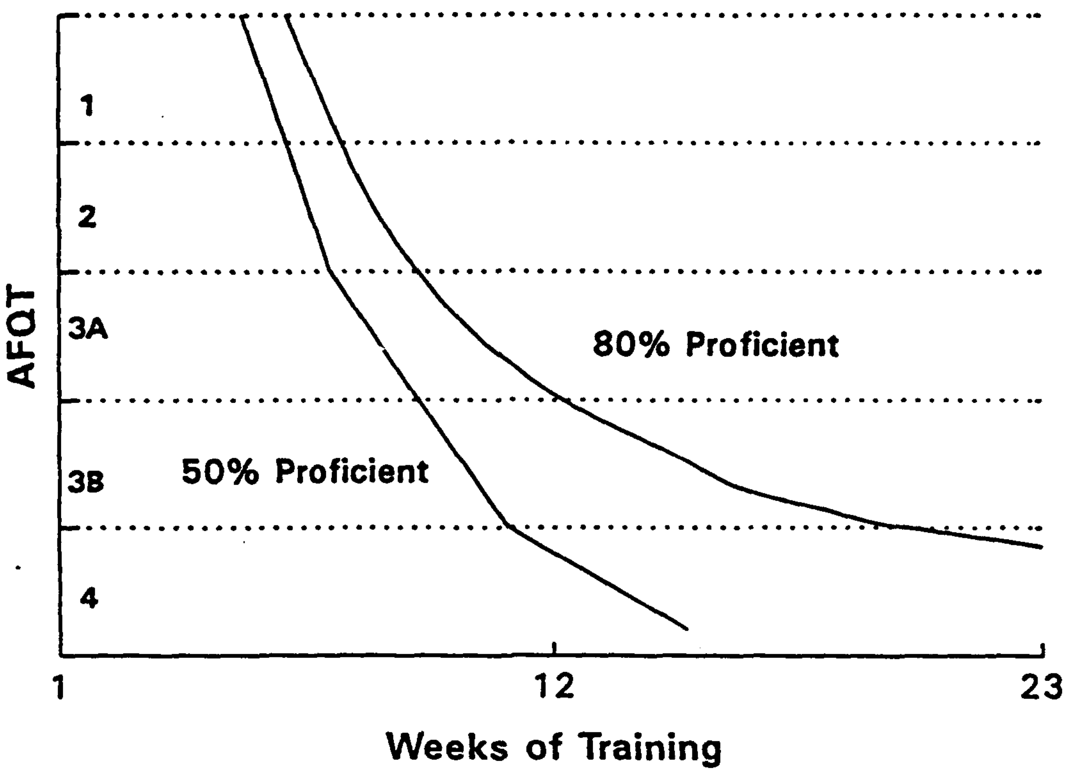

Figure 2. Two isoperformance curves, one for 80% and the other for 50% proficient.

Figure 3 presents a pair of curves quite similar to those in Figure 2, but with a very different meaning. Suppose one were to automate part of an engine mechanic's job by providing him/her with more advanced computer equipment that was itself easy to use. With the new equipment, the job becomes simpler so that the same objective results can now be achieved by lower aptitude people (e.g., lower in a specific technological aptitude) or with less training time. Figure 3 depicts such a situation. Again, there are two curves, but this time the two curves correspond to two equipment variants and both represent the same level of performance. Any point on either curve suffices to produce personnel 80% of whom are proficient. Using the new equipment (B), however, the same people can be trained to the same level of performance (80% proficient) in less time. Or, for a given amount of training time, the same level of performance can be achieved with lower aptitude personnel.

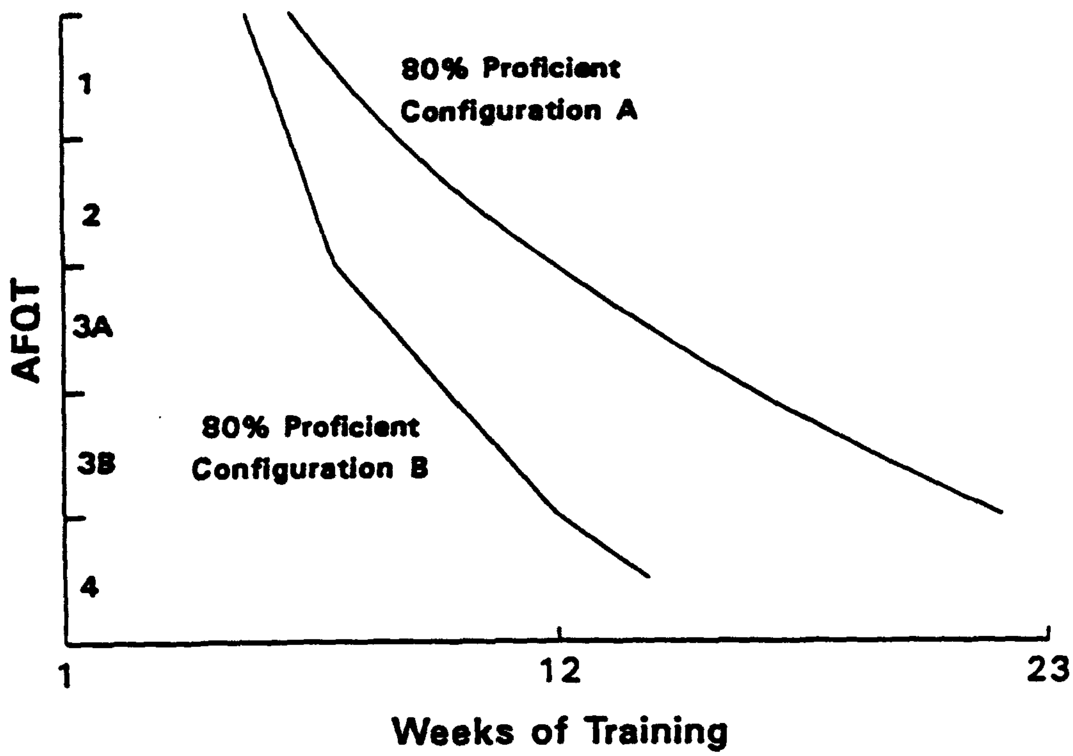

Figure 3. Two isoperformance curves, one for each of two equipment configurations, but both for the same job and the same level of performance.

In short, automating the engine mechanic's job has been successful. If the isoperformance curve for the new equipment had not shifted to the left and down, the effort to simplify the equipment and make it easier to use and learn would have failed. This sort of comparison is quite general. Any two pieces of equipment or equipment variants for performing the same job can be evaluated by comparing their respective isoperformance curves. If one curve lies to the left of and below the other, then that equipment variant is preferable to the other. Using it, the same job can be performed or learned using lower aptitude people or shorter training times.

Isoperformance curves must be evaluated before any conclusion can be reached. Any point on either of the two curves in Figure 3 will produce 80% proficient students -- but which point is best? To answer this question one invokes cost considerations. Category 1 and 2 individuals may be in such demand for other jobs that they must be regarded as unavailable. Training times in excess of 12 weeks may be excessively expensive. Figure 4 re-presents Figure 3 marked to reflect these two considerations. Since category 1 and 2 students are excluded by reason of unavailability, and category 3 students (or lower) require more than 12 weeks to reach 80% proficient using the original equipment, there is no solution to be obtained using equipment configuration A. The alternative equipment, however, does provide a range of solutions. Any point on the lower curve between the horizontal and vertical bars would be acceptable insofar as personnel

availability and training costs are concerned. They might not be equivalent, however, on other counts. It might be, for example, that training schools for engine mechanics must last at least eight weeks, shorter lengths of time being impractical for scheduling reasons. The solution would then have been narrowed to the second equipment configuration (B), category 3B and 3A individuals, and a training time between eight and twelve weeks.

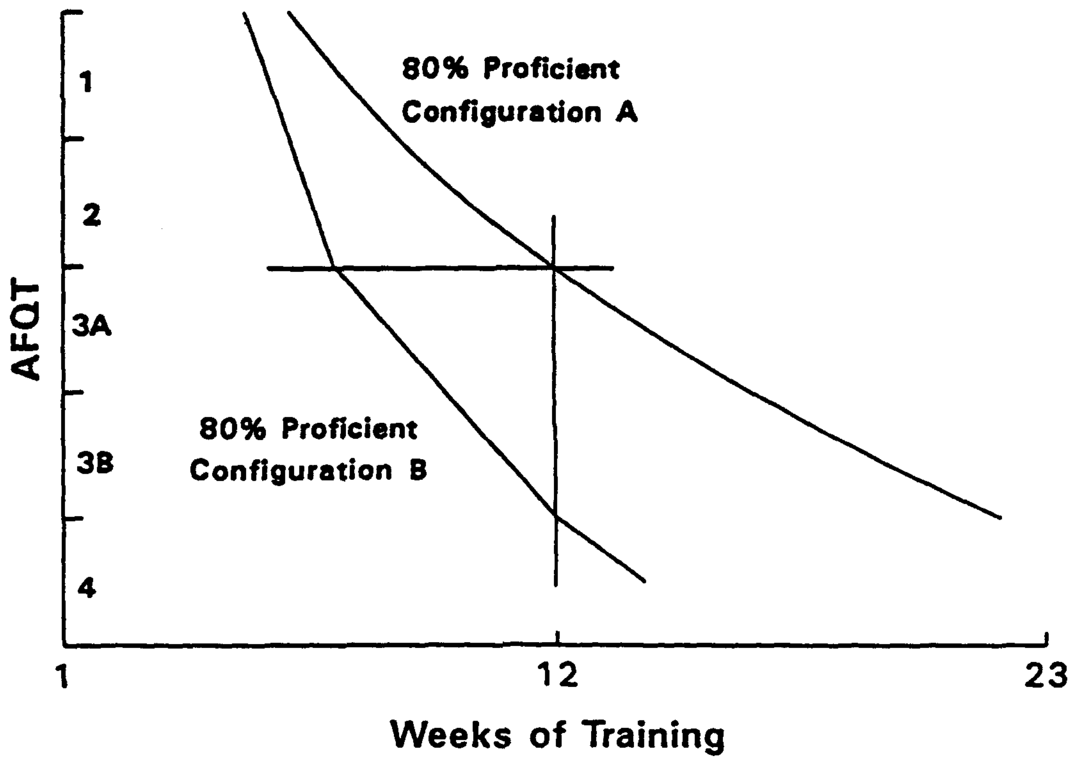

Figure 4. Figure 3 marked to indicate that category 1 and 2 individuals are not available and training times in excess of 12 weeks are too expensive.

If costs of recruitment at different aptitude levels and costs of training are known or can be estimated, a minimum dollar cost can always be determined. This point will be developed further in the next section on simulator-field trade-offs in collective training. What matters here is that a cost-benefit analysis (all relevant quantitues evaluated in dollar terms) can be conducted and will allow us to narrow the "solution space" to a point rather than an interval. If that point lies within the interval of acceptable solutions defined by considerations of personnel availability, feasible training time, safety, and the like, all is well. If not, a suboptimal solution from a purely dollar point of view will have to be accepted.

The constraints in Figure 4 could come from a HARDMAN analysis. If so, Figure 4 would then "fill in" the solution space. It would provide the trade-off analysis within constraints that HARDMAN III needs to reach a final design solution.

The same result could also be reached "the other way around," that is, by starting with the HARDMAN constraints and then carrying out the isoperformance analysis. In either case, a cost-benefit analysis is usually possible, with a consequent narrowing of the solution space to a single point.

SIMULATOR-FIELD TRADE-OFFS

When the determinants of performance are simulator training on the one hand and field training on the other, isoperformance methodology subsumes an already-established form of cost-effectiveness analysis, originally developed by Roscoe (1971). Roscoe focused on flight training, but the application to Army collective arms tactical training and CATT is direct, either as a metric to compare: (1) Distributed Interactive Simulation (DIS) with training in individual simulators, or (2) Field-Based Training exercises to DIS training, and we would expect that this relation would hold for all subsystems of CATT. In the Roscoe work, the first hour of flight simulator training typically saves more than an hour of flight training. That is, students who have experienced an hour of simulator training subsequently reach criterion performance in flight training in less time (more than an hour) than students who have had no simulator training. The second and third hours of simulator time save smaller amounts of field (flight) training time. The savings obtained from each additional hour of simulator time steadily decreases until after 14 hours in the example he used, an additional hour of simulator time produces essentially no savings in flight time. Roscoe called these savings or, better, their ratio to additional simulator time the incremental <u>transfer effectiveness ratio</u> (TER). If X denotes hours of simulator time and Y denotes hours of subsequent flight time needed to reach criterion, then

$$\text{Incremental TER} = \frac{Y_X - Y_{X + \Delta X}}{\Delta X}$$

For example, if students who have 3 hours of simulator time require an additional 6.48 hours of flight time to reach criterion, and students who have 4 hours of simulator time require an additional 5.68 hours of flight time to reach criterion, then

$Y_3 = 6.48,$

$Y_4 = 5.68,$

$X = 3,$

$\Delta X = 1,$

and incremental TER at $X = 3$,

$$\text{Incremental TER3} = \frac{6.48 - 5.68}{1} = 0.80.$$

The point is that incremental TER varies with X. It decreases, but by smaller and smaller amounts as X increases. That is, incremental TER is a decreasing, decelerating function of simulator time.

Simulator time is almost always less expensive than field training time. Hence, incremental TER has direct implications for the overall cost of training student pilots to criterion. Roscoe pointed out that if one knows incremental TER and one also knows the cost of simulator and field training, then the least expensive combination of simulator and field hours can be determined. Specifically, if r equals the cost of one hour in the simulator divided by the cost of one hour in field training, then the overall cost of training students to criterion is minimal when

Incremental TER = r.

Initially, incremental TER is greater than r and the savings in reduced field training time more than compensates for an additional hour of simulator time. Ultimately, however, the cost of an additional hour of simulator time is greater than the savings in field training. After 14 hours of simulator time in Roscoe's illustration, an additional hour of simulator time simply adds to the overall cost of training without compensation in reduced flight time. In between these two extremes, the overall cost function reaches a "crossover point" and at that point the overall cost of training is minimal. That crossover point occurs when incremental TER equals the cost ratio.

In his treatment of the trade-off between simulator and flight training, Roscoe never presented an isoperformance curve, although he did present in tabular form the data necessary to construct one. The isoperformance curve in Figure 5 was constructed from data used by Roscoe to illustrate incremental TER. It traces out all combinations of simulator time and subsequent flight time sufficient to bring the average student to criterion performance. As simulator time increases, subsequent flight time decreased, but at a descending rate. Thus, the isoperformance curve, as well as the curve for incremental TER, is decreasing and decelerated.

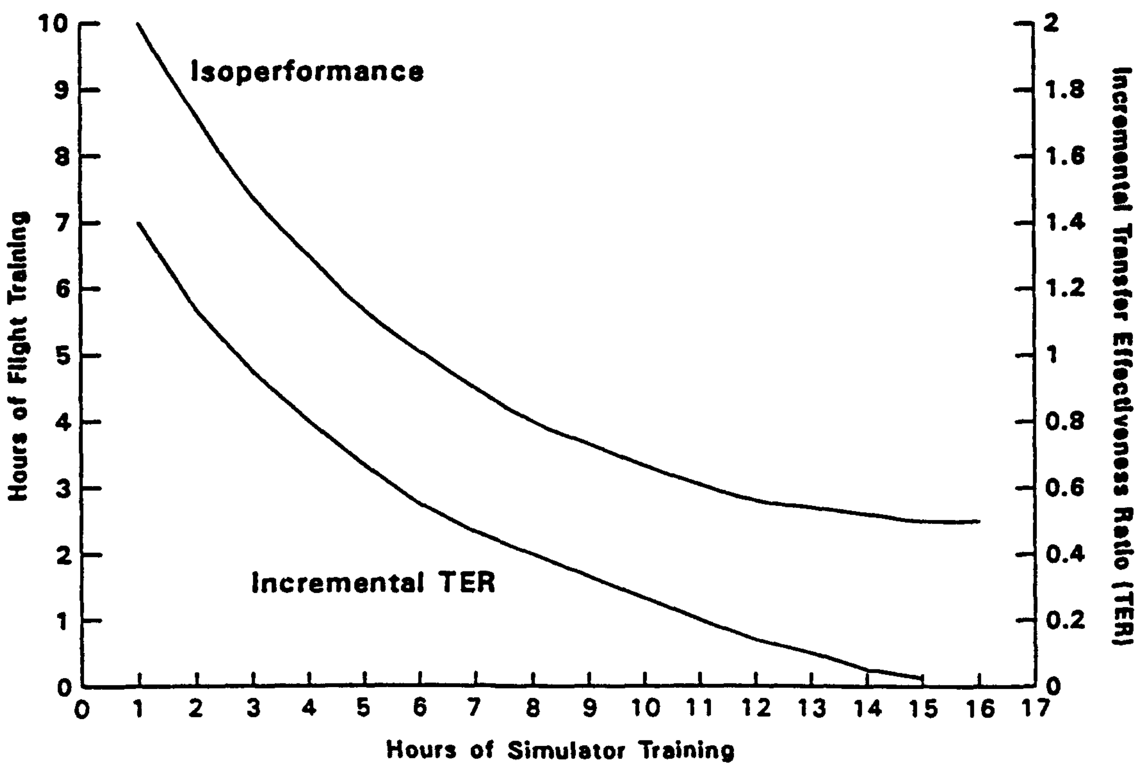

Figure 5. Transfer effectiveness ratio (TER) for the trade-off between simulator and flight training. (Data from Roscoe, 1971.)

Roscoe's incremental TER bears an obvious relation to isoperformance. Calculated in finite amounts such as an hour, incremental TER approximates the slope of the isoperformance curve; that is, it approximates the derivative of that curve with respect to simulator time. The only difference is that Roscoe takes the decrement in flight time as savings, thereby reversing the sign of the derivative. The "overall cost function," as we have been calling it,

$$\text{Cost} = c_1 S + c_2 F,$$

where c_1 is the cost of an hour of simulator time, c_2 is the cost of an hour of flight time, S is simulator hours, and F is flight (or field) hours. Taking the derivative of this function with respect to S and setting equal to zero gives us

$$\frac{d\,\text{Cost}}{d\,S} = c_1 + c_2 \frac{d\,F}{d\,S} = 0$$

or

$$\frac{d\,F}{d\,S} = -\frac{c_1}{c_2} = -r.$$

This, of course, is the same as Roscoe's result, except, again, for the change of sign.

Roscoe introduced transfer effectiveness ratios in 1971. Since then, a considerable body of literature has grown up around the idea (Orlansky, String, & Chatelier, 1982), all of which can be analyzed (or reanalyzed) in isoperformance terms, since transfer effectiveness ratios are derivative from isoperformance theory. Some of this we would propose to do in Phase II. Most of the literature has to do with flight training. This, however, is happenstance. The logic of isoperformance and of transfer effectiveness ratios holds equally well if the simulator is designed for tank or some other kind of training rather than flight training. In the remainder of the discussion, therefore, we will write about simulator-<u>field</u> trade-offs, where the skills to be acquired in the simulator or the field are not limited to aviation. In Phase II we will propose to reanalyze the Orlansky et al. data for application to the CATT family of systems.

We suggested earlier that equipment variants for performing the same job can be evaluated by comparing isoperformance curves involving personnel aptitudes and training times, and in the next section provide an example with operational data. However, equipment (or single simulation versus collective training) variants can also be evaluated by comparing isoperformance curves for simulator-field trade-offs. For example, because of cost considerations, one may wish to provide a given simulator with area-of-interest (AOI) displays rather than the more expensive wide field of view (FOV). The hazard is that the lower fidelity of AOI may not provide as effective overall simulator training, with the consequence that more field training will have to be provided. Increased field training, however, could cost more than what was saved by substituting AOI for FOV. The question is how to resolve this issue. Isoperformance provides an answer.

Suppose we have specified a level of performance at the end of field training that, as specified by the operational person responsible, is "good enough." This level could be achieved by training exclusively in the field as was the rule before simulator and training devices. It could also be achieved by some amount of simulator training and a little less field training. Moreover, collective training may be compared with training in a single component. As simulator time increases, the amount of field training necessary to reach the specified level of performance decreases, almost certainly at a decreasing rate. That is, the gain in reduced field time for a unit increase in simulator time is larger for the first hour than it is, say, for the fifth or sixth hour. Such a curve (an isoperformance curve) could be traced using AOI and another curve using FOV. Based on what we know from TER results and from other findings in the isoperformance literature (Kennedy, Jones, & Baltzley, 1989), the result would be two curves like those that appear in Figure 6.

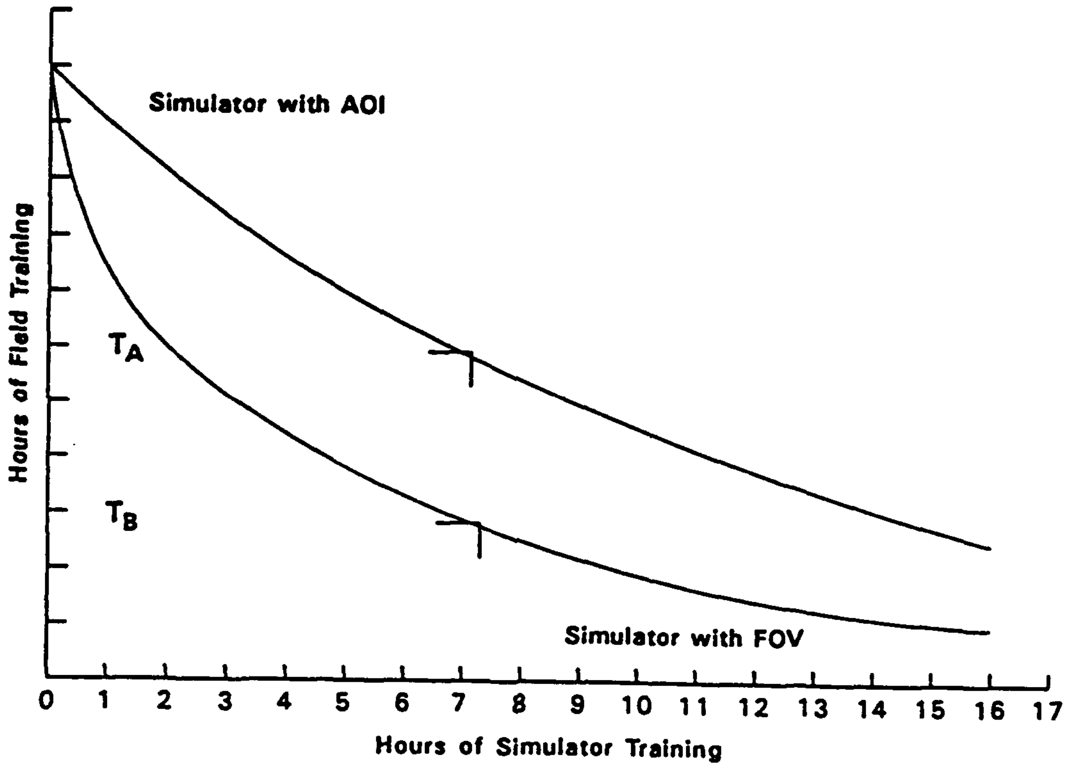

Figure 6. Hours of simulator training and hours of field training sufficient to produce a specified level of performance in the field using two simulators, one with AOI and the other with FOV.

If the AOI curve fell on top of the one for FOV, there would be no penalty in increased field time for using the AOI display, and since AOI is bound to be less costly, a decision could be made in its favor without hesitation. However, if the AOI curve fell to the left of and below the FOV curve (as depicted), the case is more complicated. Consider a given number of simulator hours, say, the point marked with an arrow (↓). In order to reach the

specified level of performance, this number of simulator hours will require T_A hours of field training, whereas the same number of simulator hours using FOV would require only T_B hours. The difference between T_A and T_B or the cost in dollars of the additional training is the penalty one pays for using AOI. If the penalty exceeds the reduced expense of AOI, then it would not be cost-effective to use AOI. On the other hand, if the penalty is less than the reduced expense of AOI, then using AOI would be cost-effective -- subject to two caveats.

First, one must take into account the number of soldiers one expects to train, that is, the total expected cost of the increased field training. This would include the number of months one expects to use the AOI simulator. The second caveat is obvious from Figure 6, namely, that the penalty paid for AOI varies with the amount of simulator training given and, therefore, with the amount of field training needed to reach the specified level of performance. Other considerations than AOI versus FOV enter at this point, for example, the availability of field training, availability of simulator training, and the overall cost function for given amounts of simulator and field training. However, given Figure 6, the role of AOI versus FOV displays could be taken into account for any stated number of simulator hours.

Isoperformance theory was originally developed for individual aptitudes and training, but it works just as well for collective training. In the CATT units as large as companies or battalions are trained at the same time. Soldiers are presumed to be skilled at their individual tasks. The purpose of CATT is to train soldiers to fight collectively in sizeable units under near-battlefield conditions. In exercises of this sort there may be dozens or even hundreds of simulators involved, all linked with one another through a common network. Here, however, just as much as in individual training, there is no thought of doing away with field training altogether. No matter how much distributed interactive simulation (DIS) is used in training, the ultimate purpose is to train the units involved to criterion in the field, just as the ultimate purpose of flight training is to train student pilots to criterion in aircraft. DIS will be used only to the extent that it is cost-effective, that is, to the extent that the savings in field time offsets the costs of additional CATT training. Simulation and field training trade-off here just as they do in individual training, and the resultant isoperformance curves and incremental TERs are as germane here as they are in individual training. The same issues arise: after how many hours of DIS training is the cost function minimized, which of two simulations or simulation variants is to be preferred, and for which levels of aptitude or prior training.

CATT will be used not only as a preliminary for the National Training Center but also for the training of combined arms and large units. In both of these cases, isoperformance trade-offs are critical within CATT training. In combined-arms exercises, the participating soldiers are first trained in "pure" arms, that is, unit exercises involving a single kind of force; for example, tanks, infantry, aircraft or artillery. Subsequently, the same men are trained in exercises involving two or more different kinds of force. Criterion performance is defined in terms of the combined-arms exercise. The question then arises as to how much training the participating units should be given in pure-arms exercises before moving on to the combined-arms exercise.

Both kinds of training are simulation-based, but formally the trade-off is similar to that between simulator and field training.

CATT will also be used to train large units; but before that can be done the participating smaller units must themselves be trained. Instead of proceeding to combined arms, the progression is from smaller to larger unit. How much time should be spent in training the smaller unit and how much when that unit is part of a much larger force? The same analysis applies here as in simulation-field or pure-combined arms, and here, just as in the earlier applications, it is critical to know the exact lie of the isoperformance curve, where cost is minimal, how isoperformance curves for different simulation, scheduling or mission variants compare with one another, and how these comparisons vary as a function of aptitude or prior training.

REAL-WORLD ISOPERFORMANCE CURVES

Related to comparisons of various training approaches (individual, collective, distributed, etc.) against the criterion of field performance, there are other applications of isoperformance to field performance which have relevance for influencing MPTF trade-off decisions. These other applications serve to emphasize the generalizability of the isoperformance model and can involve training, equipment, and manpower trade-offs.

For example, it may be different in the future, but at present no data archived by any of the Armed Services were collected with a view to isoperformance analysis. As a consequence, any attempt to construct isoperformance curves from archived data must be opportunistic. Archived data will certainly not be organized, for example, into groups each of which is given different amounts of training time and all of which are evaluated in the same terms. If the data allow the construction of isoperformance curves at all (something that is far from certain), the curves will have to be "constructed."

TPDC maintains several files of training and performance information. One of them is the "Unified SSN" file which contains all ASVAB scores, the service primary occupation, and much else. Another is the "Army SQT" file, which contains all scores on the Skill Qualification Test (SQT), the MOS and Skill Level in which a particular SQT is given, and other information. These two files were merged by Social Secrity Number so that aptitude information (ASVAB) and performance information (SQT) were brought together in the same file.

The Skill Qualification Test is given annually to soldiers in their nits. It is a paper-and-pencil test of proficiency in a Skill Level (1-5) of an MOS. One would have preferred a performance test but the only performance test data archived at TPDC are from Project A, are limited in scope and number, and do not involve repeat information (which isoperformance analysis requires). At what is now TPDC, the SQT, on the other hand, is available from most MOSs, for many in appreciable number and is repeated (in parallel forms, of course) each year. It is recognized by the Army as the best performance measure available.

The general outline of the analysis we conducted for the Phase I project was to determine the feasibility of trade-offs between aptitude as indicated by the ASVAB and time in unit at Skill Level 1, with performance on the SQT as the criterion. SQT-scores, as will be seen, improve with time in unit. The first time a soldier takes the SQT he has been in the unit a little more than six months. A year later he has become more proficient and his score is higher. If he stays another year, it is higher yet. These increases are quite what one would expect. In part, the soldier is being trained on the job and, in part, he is acquiring experience. On both counts proficiency improves.

In our analysis of this data, it remained to specify the particular MOSs for which the analysis would be conducted. Here the choice was wide. It was narrowed, however, by several considerations. First, we wanted an Army MOS. Second, the MOSs chosen had to contain a sufficient number of soldiers in recent years (1985-1991). Second, it seemed preferable to choose MOSs for comparable pieces of equipment so that the role of equipment could be illustrated. Finally, we wanted something that we could relate to the Close Combat Tactical Trainer. For these reasons we decided on four MOS turret and system mechanics for the M60 and M1 tanks. The process of constructing the isoperformance curve will be illustrated for turret mechanics on the M1 tank.

The Aptitude Measure

The first step in the analysis was to decide on the most appropriate aptitude measure. Since performance on the SQT is the criterion, a reasonable choice would seem to be that one of the ASVAB tests or composites that best predicts SQT-1 (performance on the soldier's first SQT). Soldiers, it should be pointed out, enter their units at Skill Level 1 for their MOS. A year later most of them have advanced to Skill Level 2 and take another SQT. A year after that even fewer soldiers remain at Skill Level 1. This trend toward smaller samples at SQT-2 and SQT-3 poses obvious problems not only for sample size but also for selective retention at Skill Level 1. By and large, the soldiers who remain at Skill Level 1 for two or three years are not as able and did not perform as well on SQT-1 as other soldiers. This problem will be dealt with shortly. At this point its relevance is to explain the use of correlations with SQT-1 as the basis for deciding on the most appropriate aptitude measure. The later SQT tests are much smaller in sample size, biased toward poor performance, and restricted in range.

Table 1 contains correlations between the ASVAB tests and composites and SQT-1 for the 433 Level 1 turret mechanics on the M1 tank for whom data were available on all measures. As can be seen, the correlations for the ASVAB tests are decidely lower than for the composites. Among the latter Comp-04, General Maintenance, has the highest correlation with SQT-1, .3844, while Comp-03, Mechanical Maintenance, has the next highest correlation, .3674. The latter composite, 03, is the one indicated by the Army as the qualifying area for this MOS. Hence, either it or Comp-04 could have been used in the analysis. We chose 04, primarily on the grounds that its validity in this sample is slightly higher than that of Comp-03.

Table 1. Correlation Between ASVAB Test and Composite Scores and Score on SQT-1, Level 1, for Turret Mechanics on the M1 Tank in a Sample of 433 Entry-Level Soldiers*

ASVAB Test	r	ASVAB Composite	r
01	.2921	AFQT	.2640
02	.2347	01	.3180
03	.2844	02	.2227
04	.1569	03	.3674
05	-.1067	04	.3844
06	-.0514	05	.2558
07	-.3410	06	.2956
08	.1699	07	.3330
09	.2542	08	-.2227
10	.3045	09	.3165
11	.2786	10	.3358

*Two soldiers who had scores on Comp-04 were lacking scores for one or more other ASVAB measures. Hence, sample size for SQT-1 was 435.

The next step was to divide the soldiers as nearly as possible into upper, middle, and lower thirds based on their aptitude score (Comp-04). Thirds were chosen because they seemed the best compromise of three considerations: first, that the categories be of equal size; second, that the isoperformance curve be based on at least three points; and, third, that the number of soldiers in a category at SQT-1 be substantial. The three categories were: Comp-04 $\geq$ 112 (High), 101-111 (Middle), and $\leq$ 100 (Low).

The Performance Figure

The performance figure (see Figure 7) displays the course of the three aptitude categories over SQT-1, 2, and 3. A great many such figures are possible. Figure 6 plots the medians of the three categories. Figure 8 plots the 10th percentiles. That is, 90% of an aptitude category scores above the value plotted for it in Figure 8.

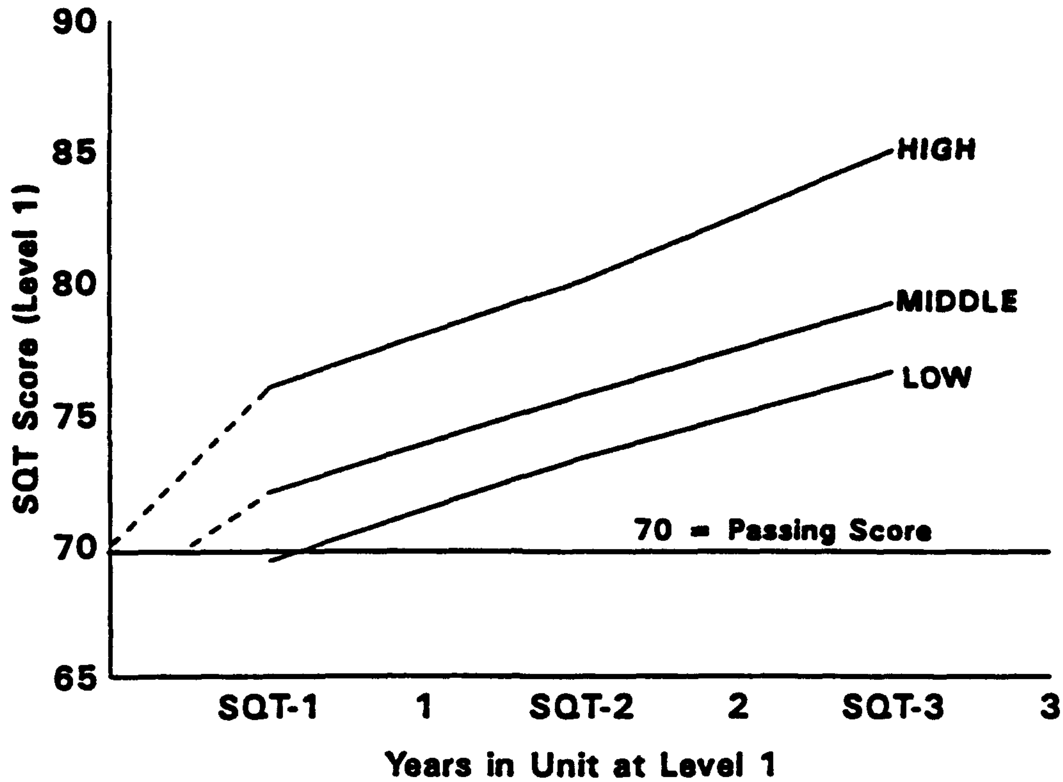

Figure 7. Median performance for the three aptitude groups at SQT-1, SSQT-2, and SQT-3.

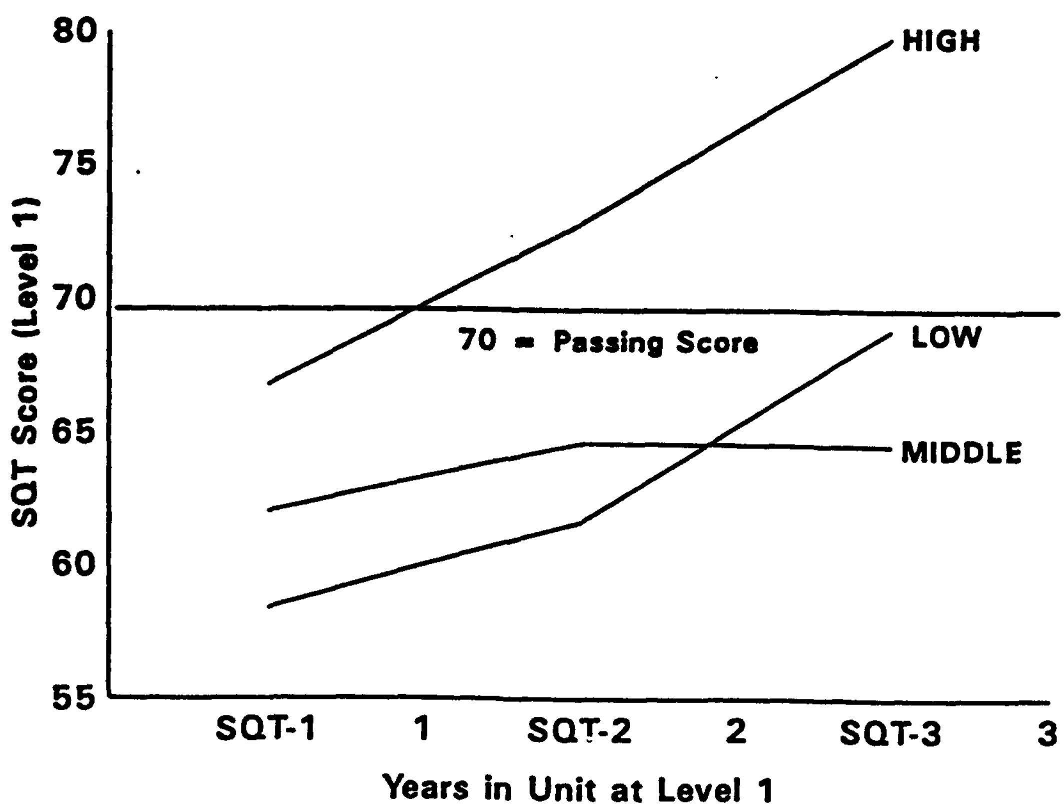

Figure 8. Tenth-percentile performance for the three aptitude groups at SQT-1, SQT-2, and SQT-3.

The next step is to explain exactly how these curves were obtained. The value for SQT-1 present no difficulties. One simply obtains the median value in each category at SQT-1. SQT-2 and SQT-3, however, do present problems. Consider the High category in Table 2. Of the 146 soldiers with SQT-1 scores, only 61 have SQT-2 scores. The remainder have progressed to Skill Level 2 before taking their second SQT at Skill Level 1. As one would expect, those who remain score somewhat lower on SQT-1, 74.0, than the general median on SQT-1, 76.0. Soldiers who are still at Level 1 in their third year (SQT-3) score lower yet, 70.5. This tendency for soldiers to score lower on SQT-1 the longer they remain at Level 1 is consistent in all three categories. Note, too, that the <u>number</u> of Low aptitude soldiers who remain until SQT-3 is twice the number of Middle or High aptitude soldiers who remain that long. Finally, note that SQT scores increase from SQT-1 to SQT-2 and from SQT-2 to SQT-3. There is only one small inversion, from SQT-2 to SQT-3 in the Middle category; and it needs to be remembered that these figures are based on only 10 cases. This tendency for SQT to increase with time in unit is also what one would expect. The possibility of a practice effect cannot be excluded but is probably small. In different years different forms of the SQT are used and a full year intervenes testings. On balance, it seems reasonable to attribute between the increase in SQT score with years in unit, as was done earlier, partly to on-the-job training and partly to increased experience.

Let us return to the High category in Table 2 and the 61 soldiers who had SQT-2 scores as well as SQT-1 scores. These soldiers gained 4.0 points between the two SQTs. It does not follow, of course, that therefore the 85 soldiers in the High category who did not take SQT-2 would also have gained 4.0 points if they had remained at Level 1 -- but it's a reasonable approximation. To be sure, one might expect soldiers who scored initially well below the mean to increase somewhat more than soldiers who scored at or above the mean. In a more refined treatment this kind of consideration might be taken into account, as well as its interaction with aptitude level. In this first report, however, we simply supposed that the rest of the High category would have gained as much as that minority of soldiers in it who remained to take both tests. On that assumption the High category would have increased from a median of 76.0 at SQT-1 to a median of 80.0 at SQT-2.

At SQT-3, the situation is similar, except that only 10 soldiers remained to take SQT-3. These 10 soldiers, however, gained 9.0 points over their performance at SQT-1. Making the same assumptions as before, the median for the High category at SQT-3 is located at 76.0 plus 9.0 or 85.0. Applying the same reasoning to the Middle and Low aptitude categories, one obtains values of 72.0, 75.7, and 79.2 (Middle) and 69.4, 73.3, and 76.6 (Low).

Table 2. Medians by Aptitude Category at SQT-1, SQT-2, and SQT-3 for Level 1 Turret Mechanics on the M1 Tank

Aptitude Category	(N) SQT	SQTs at Level 1 1	1 and 2	1, 2, and 3
High	(N)	(146)	(61)	(10)
	1	76.0	74.0	70.5
	2		78.0	76.0
	3			79.5
Middle	(N)	(140)	(62)	(10)
	1	72.2	72.0	68.0
	2		75.5	75.5
	3			75.0
Low	(N)	(149)	(65)	(20)
	1	69.4	68.4	68.0
	2		72.3	71.0
				75.2

It is our understanding that the Army regards a score of 70 on the SQT as passing. The Low-aptitude median crosses SQT=70 at 0.75 years in unit, that is, after the SQT is generally taken. The upper two groups, however, cross the pass line <u>before</u> their first taking of the SQT. In Figure 7, 50% of the High-aptitude people have been supposed to arrive at their units proficient. This supposition is not entirely hypothetical. Subject-matter experts whom we have consulted on this matter say that, if anything, this supposition is conservative; that is, at least 50% of the top aptitude group would be proficient on arrival at their units. Since the curve for the Middle-aptitude group is elsewhere roughly parallel to that for the High-aptitude group, it was extended back prior to SQT-1 by making it parallel there too to the curve for the top group.

Figure 8 was constructed in the same way as Figure 7 from the data presented in Table 3. The only difference was that 10th-percentile scores were used instead of medians. Thus, 90% of the High-aptitude group, for example, scored above 66.8 on SQT-1. Since all three groups scored below 70 at SQT-1, there was no need in Figure 8 to extrapolate back prior to SQT-1.

Table 3. Tenth Percentiles by Aptitude Category of SQT-1, SQT-2, and SQT-3 for Level 1 Turret Mechanics on the M1 Tank.

Aptitude Category	(N) SQT	SQTs at Level 1: 1	1 and 2	1, 2, and 3
High	(N)	(146)	(61)	(10)
	1	66.8	64.6	59.0
	2		70.6	71.5
	3			72.0
Middle	(N)	(140)	(62)	(10)
	1	62.0	62.1	62.0
	2		64.7	65.5
	3			64.5
Low	(N)	(149)	(65)	(20)
	1	58.4	58.2	57.5
	2		61.4	61.5
				68.0

The feature of most interest in Figure 8 is what happens with the Middle and Low groups. The Army rule of thumb is 90% proficient (Mayberry, 1992); any figure less than that is thought to jeopardize readiness. Yet, in Figure 8, the bottom two thirds of the tank turret mechanics do not reach 90% proficient by their third SQT. More than 10% of the bottom third (20 out of 149) are still at Level 1. If our construction is even approximately correct, however, in the Middle third some soldiers must have been advanced to Level 2, even though they did not have passing marks on their most recent SQT.

Isoperformance Curves

Any performance figure yields an isoperformance curve. In Figure 9, for example, 50% of the High-aptitude people are proficient on arrival at their units; the Middle-aptitude group reaches 50% proficiency at 0.35 years, and the Low-aptitude group at 0.75 years in units. If the categories are plotted at their median percentiles (83rd for the High, 50th for the Middle, and 17th for the Low group), one obtains the isoperformance curve labeled M1 in Figure 9. The 90%-proficient isoperformance curve for turret mechanics consists, it will be recalled, of a single point for the High group, because neither of the lower two groups reach 90% proficiency by SQT-3.

The second curve in Figure 9 is for turret mechanics on the M60 tank. It was obtained using the same procedures as were used to obtain the M1 curve. These two curves differ, of course, in the equipment it is the mechanic's job to maintain. They also differ, however, in three other important ways. First, the two curves involve different groups of soldiers, and such a difference can affect the lie of the isoperformance curve. A tendency, for

example, to assign abler soldiers to new equipment would shift the performance curves for all three aptitude groups in Figure 9 up and to the left. This shift, in its turn, would shift the isoperformance curves down and to the left. It is possible, therefore, that the M1 curve in Figure 9 lies to the left of the M60 curve for this reason. It is unlikely, however. Figure 10 presents the isoperformance curves for systems mechanics on the two tanks, and there the M60 curve lies to the left of the M1 curve, just the reverse of Figure 9. There seems no obvious reason why above-average turret mechanics and below-average system mechanics would be assigned to the M1 tank. Furthermore, the differences between the two curves in Figures 9 and 10 are small and certainly not significant. The larger fact in both figures is not difference but communality.

A second difference between the curves in Figure 9 concerns the aptitude measure. For the turret mechanics the best-predicting ASVAB composite for the M1 tank was Comp-04, General Maintenance. For the M60 tank it was Comp-09, Skilled Technical. A difference in predictors could also affect the isoperformance curve. One predictor, for example, might be substantially more valid than the other. The effect of increased validity is to make the isoperformance curve more aptitude-sensitive, that is, less steep, so that it trails out more to the right. The difference in validity between the predictors for the M60 and M1 tanks was small, and what difference there was favored Comp-04, which would produce the opposite effect from what appears in Figure 9. If a difference in validity were at work, the M1 curve would lie to the right of the M60 curve.

Still a third important difference is the SQT itself. If a task is automated with a view to making it easier to operate, then the corresponding SQTs will change. In a sense they will become "easier" but in a way that reflects the changes in the equipment and does not represent a lower standard of effectiveness either for the tank as a whole or for any of its crew members. Such a change would produce a true, positive shift (down and to the left) of the isoperformance curve. It would represent a real reduction in the personnel and training requirements of the tank.

It is also possible, however, for the SQT to become easier in a spurious way by simply lowering standards of effectiveness. The effect would be the same as using a less stringent criterion, for example, 70% proficient instead of 90% proficient. The isoperformance curve would shift down and to the left. There was insufficient time in the Phase I contract to explore this possibility in any depth, but the reversal in Figure 10, making the M60 tank easier to maintain than the M1, clearly indicates that there were no general lowering of standards in maintenance for the M1.

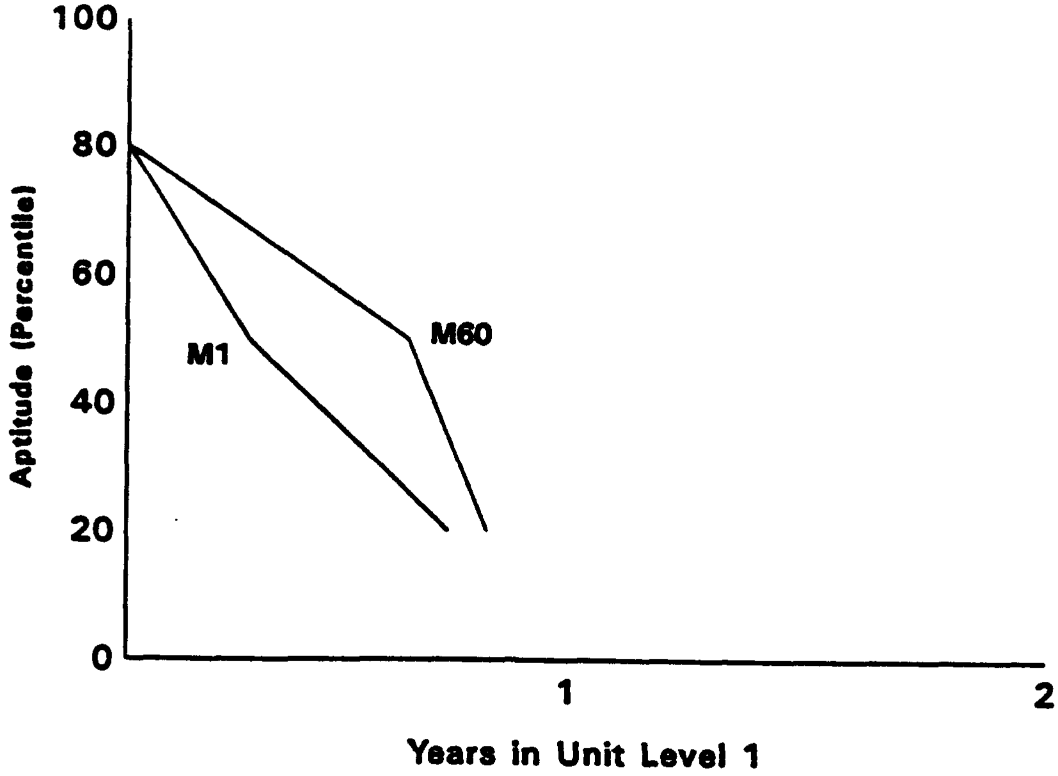

Figure 9. Isoperformance curves (50%) for turret mechanics in the M60 and M1 tanks.

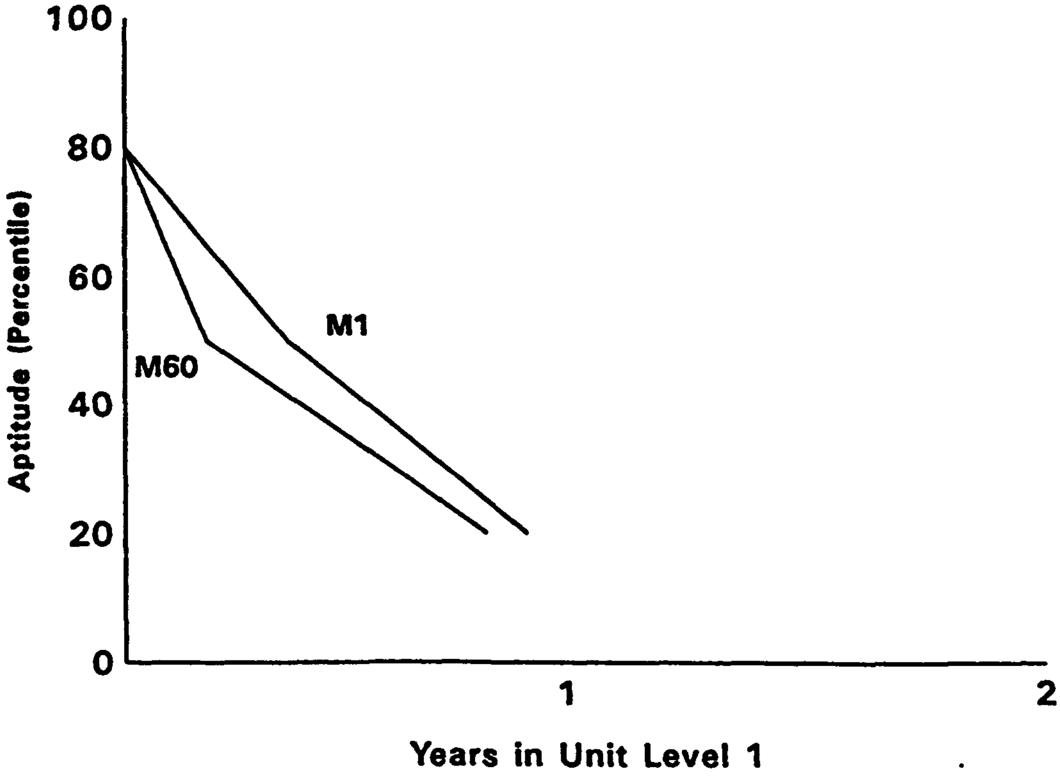

Figure 10. Isoperformance curves (50%) for systems mechanics in the M60 and M1 tanks.

In fact, there is probably no reliable difference between the two curves in Figures 9 and 10. The four curves are all very similar. The High-aptitude group in all four cases arrive at their units 50% proficient. This, of course, is largely artifactual since all four points were fixed by supposition. On the other hand, the medians of all four High-aptitude groups were quite close to each other at SQT-1, varying from 74 to 78. Clearly, all four isoperformance curves began earlier than when the SQT-1 was typically administered and quite possibly close to each other around the time of arrival at the units.

The four Low-aptitude groups are all fixed mainly by TPDC data and are also located close to each other, between 0.75 and 0.95 years. The Middle-aptitude points fall in between, possibly a little earlier for the system than for the turret mechanics.

So altogether the 50% isoperformance curves for the four MOSs are very much alike. The High-aptitude mechanics arrive at their units close to 50% proficient. The Low-aptitude mechanics do not reach 50% proficiency until late in their first year. And the Middle-aptitude mechanics reach 50% proficiency after 4 or 5 months.

Cautionary Comment

Sample size for the system mechanics was much smaller than for the turret mechanics, 146 instead of 435. As a consequence, fewer than five soldiers remained at Level 1 long enough to take SQT-3. Hence, the performance curves for both M60 and M1 tanks stop at SQT-2. In addition, the 50% points for the High groups were fixed largely by supposition and not extensively checked. It would be premature, therefore, to regard the conclusions we have reached as definitive. At the least, however, the isoperformance curves constructed in this report constitute a consistent and, if true, informative account of the trade-off between aptitude and training-plus-time-in-unit in tank-maintenance mechanics.

BACKGROUND

Since the 1950s, applied behavioral scientists working in the fields of systems, training, and selection have remained largely independent from each other. Within any bureaucracy, organizational charts, policy management guides, and function statements can reinforce this separation. Historically, in systems research work, human factors practitioners have been taught that their role is to gather these human input/output data (transfer functions) of humans and determine how they interact with their equipment (or physical and environmental stimuli). These data would be used to generate standards and specifications which could then be used by design engineers to improve systems performance. They have also been led to believe that design engineers were eagerly awaiting those data to incorporate into new systems which would permit efficient allocation of functions between man and machines (Fitts, 1963; Taylor, 1963). This goal, while lofty, was naive and one of the intentions of the present work is to call attention to a technique whose goal is to improve decision-making in systems research by employing as a strategy the notion of "trade-off technology."

In recent years, weapons systems and products have become more advanced as technology advances and logistics requirements have increased in kind. Simulator and weapons system development contractors are faced with a dilemma because the costs related to system/product acquisition and support are increasing at alarming rates, while the climate of decreasing military budgets results in less money. In view of these trends, one of the greatest challenges facing the weapon system acquisition process (WSAP) today is to meet the growing need for more effective and efficient management of our resources. The national push to increase productivity in an environment of tight resources has placed emphasis on all phases of the simulator weapon systems life cycle. As a result, one primary requirement for weapon systems developers is to analyze the logistics, human factors, and MPTS implications of alternative approaches as part of the weapon systems design process in order to maximize human systems effectiveness.

Some of these human systems are identified in logistic support analysis (LSA). For example, logistic support analysis is designed to incorporate human factors (HF) considerations to assure complete compatibility between the system physical and functional design features and the human element in the operation, maintenance, and support of the system (Blanchard, 1986). Considerations in design must be given to anthropometric factors, human sensory factors, physiological factors, psychological factors, and their interrelationships. In addition, major sources of manpower, personnel, training, and safety (MPTS) data must be considered early in system development.

Unfortunately, most information that we currently have concerning the human system is not well organized or easy to find, thus providing weapons systems designers with little tangible assistance in meeting military requirements and advancing productivity goals. What is more, weapon systems designers do not have the expertise nor the tools that are required to make MPTS trade-off decisions. Integrated logistic support (ILS) requires a computerized technology for aiding such decisions.

Isoperformance is a trade-off technology. Its overall purpose is to identify combinations of determinants which have the same outcome for performance. This same overall purpose has been the focus of several large-scale efforts to develop integration methodologies (e.g., MANPRINT, RAMPARTS, IMPACTS, HARDMAN). To be maximally effective, a weapon system must be operable by the crew specified for it, personnel of the required aptitude levels must be available, and the amounts of time required to train these personnel must not be so long as to be exorbitantly expensive (Department of the Army, 1986). Generally, with humans in the loop, advice about how to enhance the effectiveness of weapons systems is through training or selection or equipment elements, depending on the background of the person giving the advice and the mission statement of his or her parent organization. In practice, however, these three kinds of considerations are almost always combined through trade-offs and an implicit cost-effectiveness model. For example, if the equipment were partly automated, it might be operable by lower aptitude personnel or a smaller crew. As aptitude level goes down, training times must be lengthened to compensate. Below certain aptitude levels, no amount of training may suffice to make a necessary percentage of students proficient.

In earlier work, Essex has developed a trade-off or integration methodology based on isoperformance curves for dealing with exactly these manpower, personnel, and training issues (Essex Corporation, 1986; Jones, Kennedy, & Kuntz, 1987; Jones, Kennedy, Kuntz, & Baltzley, 1987; Kennedy, Jones, & Baltzley, 1988, 1989). By "isoperformance" we mean to identify a specified standard of human performance that can be achieved by incorporating human factors and MPTS data into weapon system design. "Iso" means "same," and isoperformance curves reflect trade-offs among human factors and MPTS variables that culminate in the same specified performance standard. The present report refocuses this work primarily in relation to training, especially collective training by means of distributed interactive simulation.

REFERENCES

Blanchard, B. S. (1986). Logistics engineering and management (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Department of the Army. (1986). MANPRINT source selection criteria: Draft guide. Washington, DC: Department of the Army, DAPE-ZAM.

Essex Corporation. (1986, April). Regions of isoperformance: Making trade-off decisions in equipment design (SBIR Phase II Proposal, Topic No. 204, Air Force). Orlando, FL: Essex Corporation.

Fitts, P. M. (Ed.) (1951). Human engineering for an effective air investigation and traffic control system. Washington, DC: National Research Council.

Jones & Jones Research Associates & Essex Corporation. (1987). A brief introduction to isoperformance methodology (computer program). Orlando, FL: Essex Corporation (Armstrong Aerospace Medical Lab Contract No. F33615-86-C-0553).

Jones, M. B., Kennedy, R. S., & Kuntz, L. A. (1987). Isoperformance: Integrating personnel and training factors into equipment design. Paper presented at the 2nd International Conference on Human-Computer Interaction, Honolulu, HI.

Jones, M. B., Kennedy, R. S., Kuntz, L. A., & Baltzley, D. R. (1987, October). Isoperformance: Trading off selection, training, and equipment variations to maintain the same level of systems performance. Proceedings of the 31st Annual Meeting of the Human Factors Society (pp. 634-637). Santa Monica, CA: Human Factors Society.

Kennedy, R. S., Jones, M. B., & Baltzley, D. R. (1988). Empirical demonstration of isoperformance methodology preparatory to development of an interactive expert computerized decision aid (Final Rep.). Washington, DC: Army Research Institute.

Kennedy, R. S., Jones, M. B., & Baltzley, D. R. (1989). Regions of isoperformance: An interactive software package for trading off training, personnel, and equipment (AAMRL-TR-89-003). Wright-Patterson Air Force Base, OH: Armstrong Aerospace Medical Research Laboratory.

Orlansky, J., String, J., & Chatelier, P. R. (1982). The cost-effectiveness of military training. Proceedings of the Interservice/Industry Training Equipment Conference (pp. 97-109). Washington, DC: National Security Industrial Association.

Mayberry, P. W. (1992). Evaluating minimum aptitude standards. Military Psychology, 4, 1-16.

Roscoe, S. N. (1971). Incremental transfer effectiveness. Human Factors, 13, 561-567.

Taylor, F. V. (1963). Human engineering and psychology. In S. Koch (Ed.), Psychology: A study of a science, Vol. 5 (pp. 831-907). New York: McGraw-Hill.